Members of the DL Lab

University of Cape Town
Department of Computer Science
Digital Libraries Laboratory @ Centre for ICT4D

May 2017

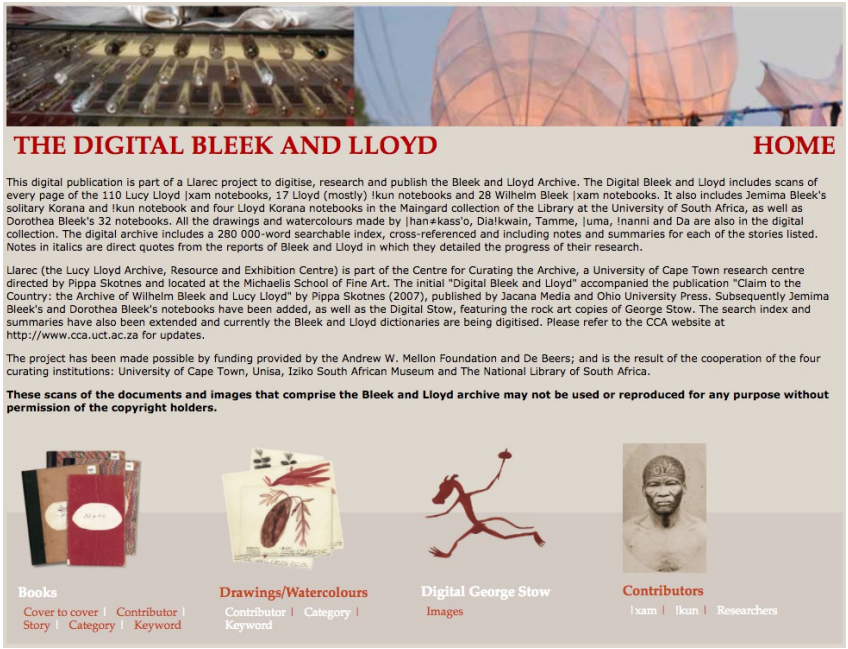
Digital Libraries

- Making information available to people
- Context-sensitive: low resource environment, different languages, different skills/culture, etc.
- 4 areas of interest:
  - Information Retrieval - HS
  - Educational Technology - HS
  - Knowledge Engineering - CMK
  - ICT4D Applications - various
- Some production systems / products

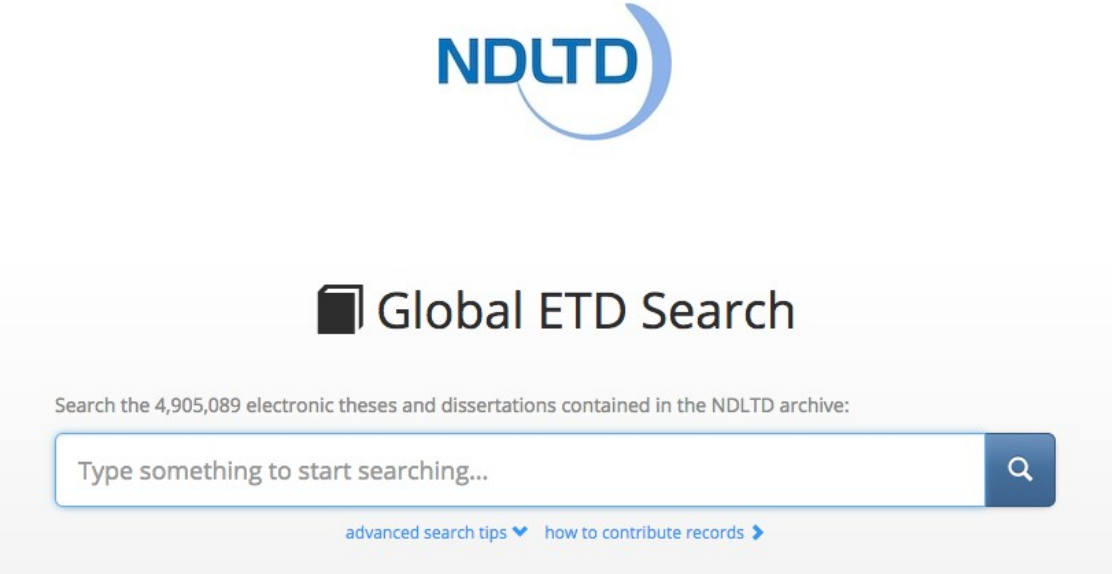
Information Retrieval and Educational Technology

- Information Retrieval (IR):
  - Search engine algorithms in African languages
  - IR for Development; Innovations in IR
  - Corpus development for African language content
  - Applied into local Digital Library applications, such as heritage archives
- Educational Technology:
  - Low-cost innovation, mobile tools
  - Focus on the teacher and not just the student
  - Modern technology and controlled studies

Information Retrieval Products



http://lloydbleekcollection.cs.uct.ac.za/



http://search.ndltd.org/

KnowledgE ENgineering team

- **Aim**: to contribute computing theory, methods, and techniques to the knowledge society
- **Scope**: knowledge engineering in its broad sense; includes ontology engineering, the Semantic Web, intelligent (logic-based, ontology-driven) conceptual modelling, and natural language generation.
- Some of the current topics:
  - OE: modularity, Test-Driven Development
  - NLG: Agglutinative languages in Sub-Sahara (data and knowledge, and mainly in health and weather)
  - CDM: model interoperability, temporal modelling
- More information: http://www.meteck.org/keen/

Information Retrieval

- Catherine Chavula (PhD)
- Selvas Mwanza (PhD)
- Jivashi Nagar (PhD)
- Gina Paihama (PhD)
- Mushashu Lumpa (M.CS)
- Jackson Moji (M.CS)
- King'ori Maina (M.CS, submitted!)
- Morebodi Modise (M.CS, submitted!)

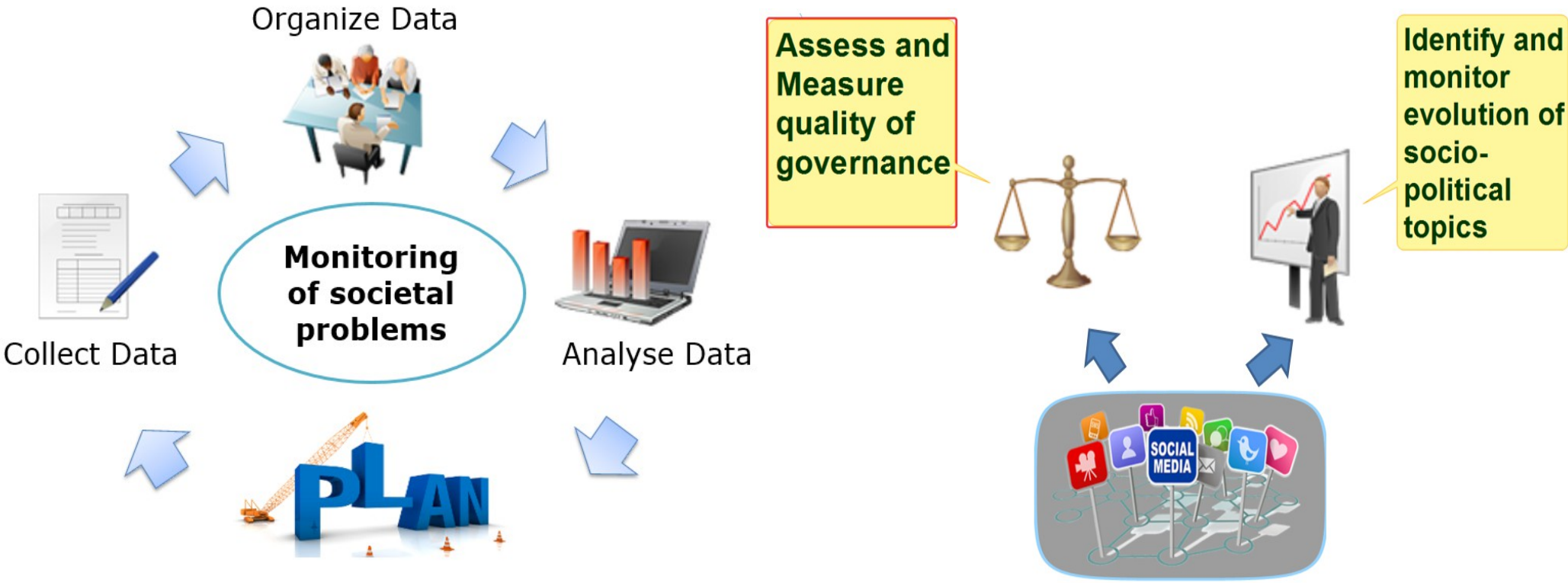
Retrieval for Related African Languages

- Context
  - Low quality search results
  - Widespread multilingualism
- Approach

Status

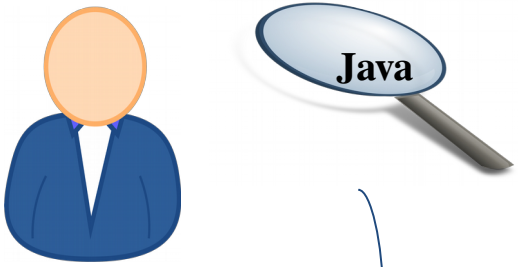
- Implementation and experiments

Social Media for Development (SM4D)

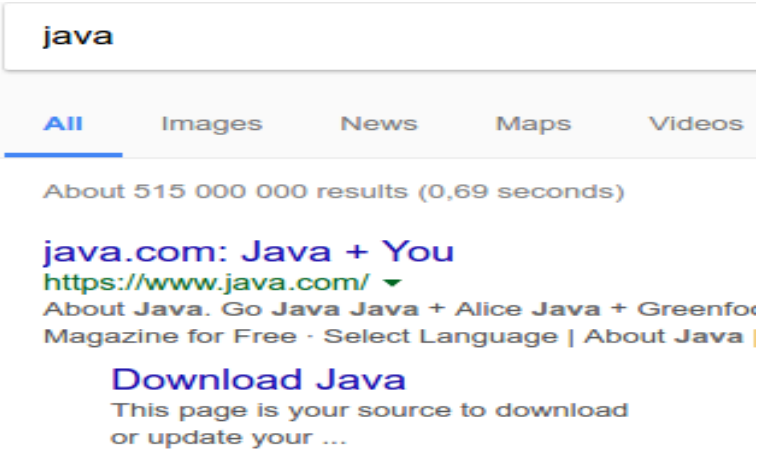


By: Selvas Mwanza

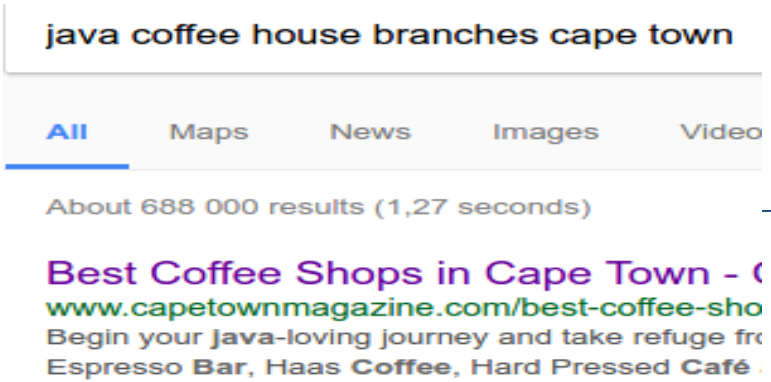
Do Topics & Time Go Together?



Actual results



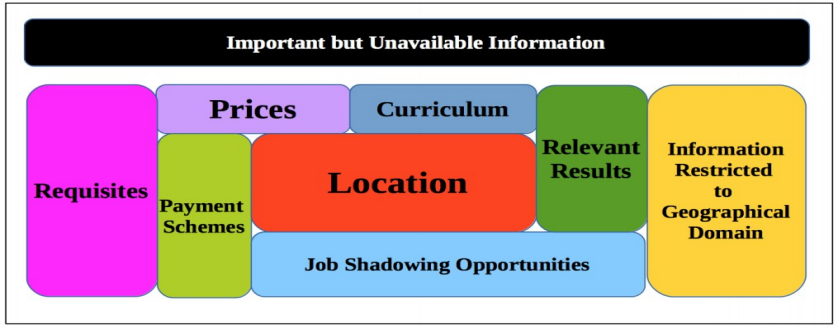
Expected results



- Topics of interest
- Time sensitive pattern
- Time of search
- Personalized reranking

Information Retrieval for Unemployment

- Key Ideas
  - Specialised IR approaches for finding training opportunities
  - E.g.: summaries, focused search
- Status
  - early experiments



- PhD candidate: Gina Paihama

Digital Libraries in Clouds

- Key Idea
  - Digital repositories increase sharing of content
  - Can we automate creating/installing/configuring in private clouds?
- Status
  - Some experiments done, some to do
- MSc student: Mushashu Lumpu

African Language Content Generation

- Problem
  - Lack of African language resources in digital form
  - Insufficient support for African languages in software
- Proposed Solution:
  - Create a focused archive of writings
  - Encourage translation of popular content e.g., Wikipedia
  - Use gamification for long-term sustainability
    - Research focus: Will this approach work?
- MSc student: Jackson Moji

Knowledge Engineering

- Joan Byamugisha (PhD)
- Zola Mahlaza (M.CS)
- Jarvis Mutakha (M.CS)
- Cameron Kyle (M.IT)
- Zubeida Khan (PhD)
- Yamiko Msosa (PhD)
- Iman Neguib (M.IT)

A Grammar Engine to Verbalize Ontologies in Runyankore

Motivation

- NLG successfully applied to generate personalized patient information

Problem

- Templates which have successfully generated personalized patient information are inapplicable to Bantu Languages

Solution: Grammar Engine

Panado ⊑ ∃ **reduces**.Pain

Direct Translation:
Panado ← mubazi gwa Panado
Reduces ← **kyendez**
Pain ← obusaasi

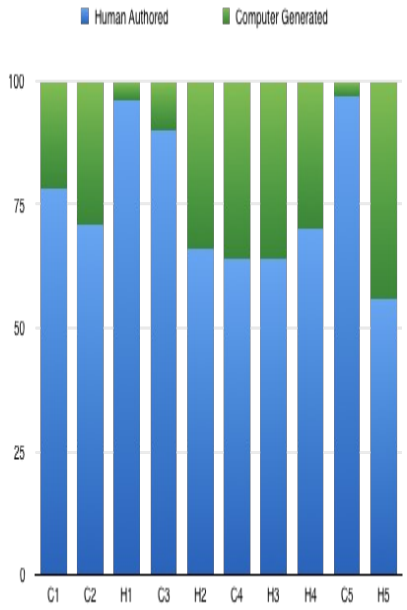
Verbalization Patterns:
⊑ ← buri
∃ ← hakiri

Noun Class System:
subject concord ← gu

Verb Conjugation:
kyendez ← nigukyendeza

Buri mubazi gwa Panado
hakiri nigukyendeza
obusaasi

Results



IsiZulu & isiXhosa weather bulletins using NLG

South African Weather Service

- There is no fast and large scale producer, automated or otherwise, of weather summaries in Nguni languages
- 'Multilingual' natural language generation system
- Our focus:
  - 'Single' grammar compatible with isiXhosa and isiZulu
  - Grammatical similarities for isiZulu and isiXhosa verbs
  - Impact of phonological conditionining in weather-based NLG (***izaukuba -> izokuba***)

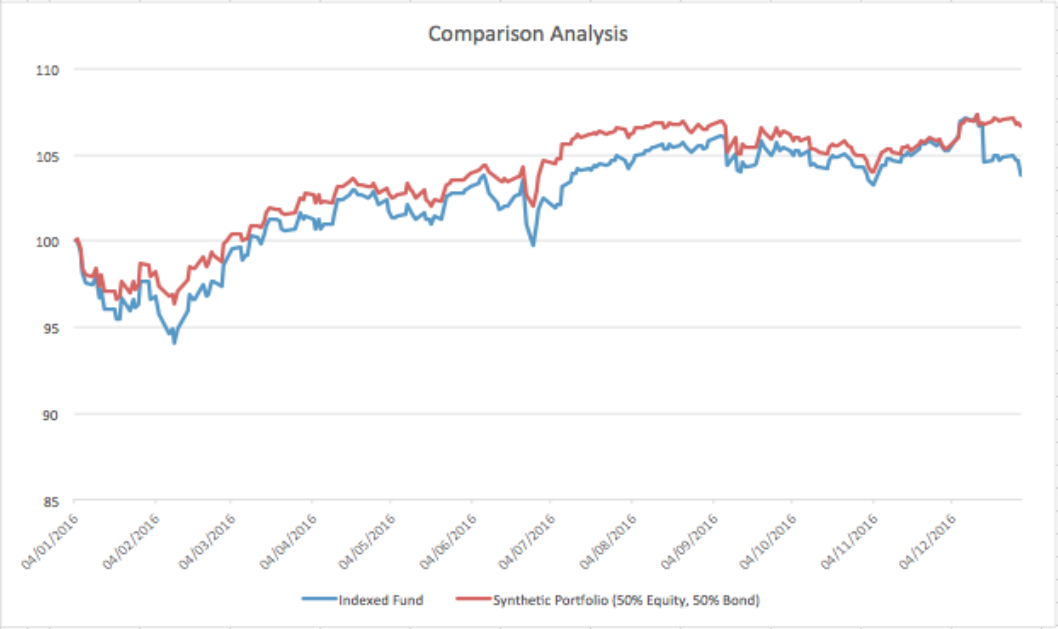
Met Office

Multiple choice question generation for learners

- Key Ideas
  - Using isiZulu and isiXhosa grammar engine
  - Generating MCQs from a domain using grammar engine
  - Measuring quality of MCQs
- Status
  - Proposal drafting
- MSc candidate: Jarvis Mutakha

NLG for financial data (Cameron Kyle)

Indexed Fund	Bond Index	Equity Index	Synthetic Portfolio (50% Equity, 50% Bond)	Daily variance	Correlation
100	100.00	100	100	0.00%	98.083%
100.07	99.97	100.2012	100.084255	0.01%	
99.36	100.28	98.88704	99.5825547	0.23%	
98.14	100.37	96.54338	98.4570656	0.32%	
97.57	100.48	95.497	97.9884016	0.43%	
97.50	100.34	95.57849	97.9609907	0.47%	
97.93	100.56	96.32427	98.4429259	0.52%	
96.72	100.74	93.91949	97.3304988	0.63%	
97.22	100.53	95.48756	98.0109412	0.82%	
96.07	100.75	93.42512	97.0887639	1.06%	
96.07	100.69	93.4748	97.0836192	1.05%	
95.43	100.80	92.38172	96.5888753	1.21%	
95.50	100.69	92.86168	96.7743345	1.33%	
96.72	100.61	94.74526	97.6752318	0.99%	
95.93	100.70	93.26364	96.9834898	1.10%	
96.65	100.79	94.58279	97.6839579	1.07%	
96.15	100.77	93.55529	97.1647563	1.06%	
96.29	100.84	94.07252	97.4560837	1.21%	
97.64	101.06	96.40177	98.7297571	1.11%	
97.64	100.89	96.35905	98.6266087	1.01%	
96.65	101.27	94.55298	97.9116787	1.31%	
96.72	101.17	95.02499	98.0958882	1.43%	
96.79	101.25	95.17007	98.2120452	1.47%	
95.72	101.28	93.41121	97.343519	1.70%	
94.79	101.64	92.08907	96.8651018	2.19%	
94.65	101.60	92.02796	96.8154628	2.29%	
94.86	101.72	92.01057	96.8667433	2.11%	
94.08	101.83	90.87873	96.3526214	2.42%	
94.93	101.42	92.65251	97.0377754	2.22%	
95.93	101.33	94.18282	97.7565859	1.90%	
96.86	101.23	95.73499	98.4836026	1.68%	
96.65	101.52	95.28832	98.4020254	1.82%	
96.65	101.53	95.28584	98.4062374	1.82%	
97.43	101.50	96.66312	99.083973	1.70%	



As shown above, there is relatively strong correlation between the funds of over 98% but a variation range of 0.21% to 2.81%, the latter occurring at the end of the period. The synthetic portfolio is generally higher than the indexed fund with a significant positive variation from July to September where the average variance was 1.33%. An event toward the end of the period appears to have caused the deviation. The average Nav was 1.07% higher for the Sythetic Portfolio compared to the Indexed Portfolio.

Ontology Engineering: Modularity (Zubeida Khan)

- Many ontologies, and large and/or expressive ones
- Problems: cognitive overload, scalability, reuse of fragments
- Solution: ontology modules
- How to create them? quality? maintainability? etc.
- Main achievements:
  - An empirically-based framework for ontology modularization
  - Evaluation metrics (and dependencies between them) for ontology modules
  - Novel algorithms for creating good modules
  - Swapping ODP modules (e.g., a foundational ontology)
  - Tooling support for these tasks
- More information: http://www.thezfiles.co.za/research.html

Changes in clinical practice guidelines (Yamiko Msosa)

- CPGs change frequently to reflect updates in medical practice
- Makes it hard to maintain software systems using CPG
- Find a way to automate that
- Current main outcomes:
  - Types of changes in the (fine-grained) components of the guidelines
  - Model-driven engineering q domain-specific language (DSL) to model, annotate the computerised CPGs
  - Evaluated DSL on adequacy and usability

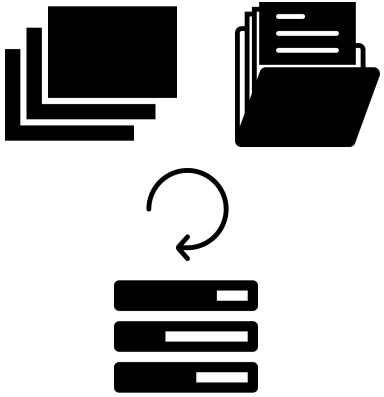
Intelligent temporal conceptual modelling (Iman Neguib)

- Temporal data and conceptual models
  - e.g., in health, business, security
    - Dev-_{PhDStudent,Academic}: "An Academic may have been a PhD student before, but is not a PhD Student now"
    - "a HIV positive patient may not evolve into a blood donor"
- How to convert a temporal model into a temporal database?
- Expected outcome:
  - Algorithms that ensure that the constraints of the temporal model apply to the actual database

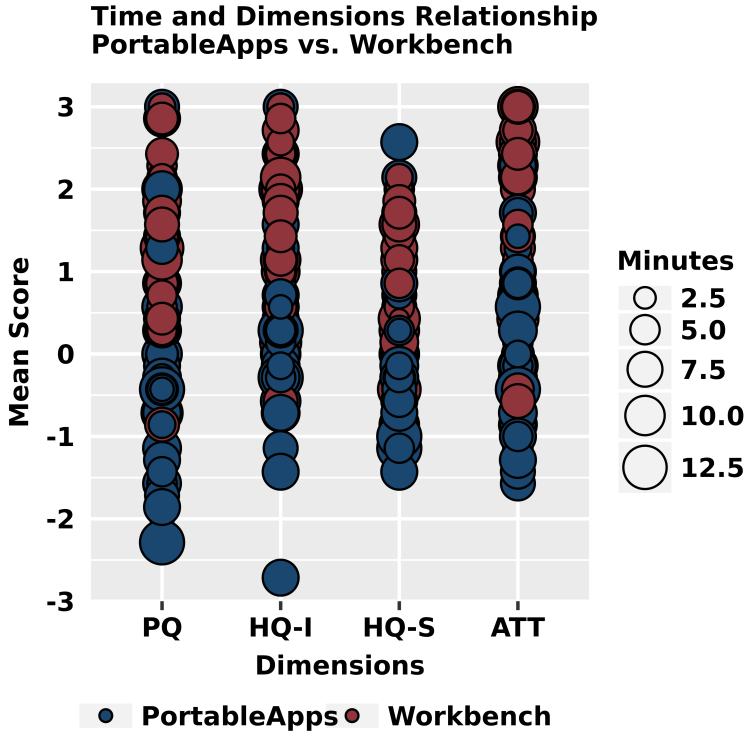
Educational Technology

- Lighton Phiri (PhD)
- Sinini Ncube (PhD)
- Maryam Almukhaylid (M.CS)
- Devan Govender (M.IT)

Streamlining Technology-driven Orchestration



1) Activity Managment
2) Resource Management
3) Sequencing
4) Playback Sequence



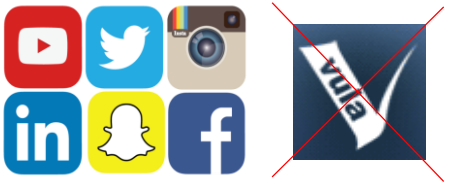
- Experimental Results
  - Effectiveness & user experience
  - Orchestration load
  - Practical usage scenarios

Integrated in-Content Communication for Shared Textbook Use

- Key Ideas
  - Mobile electronic textbooks
  - Discussion forums integrated with content
    - vs. standalone discussion forums
- Status
  - Currently writing up thesis
- PhD candidate: Sinini Ncube

A Comparison of Socially-Motivated Discussion Forums Models for LMSes

Problem

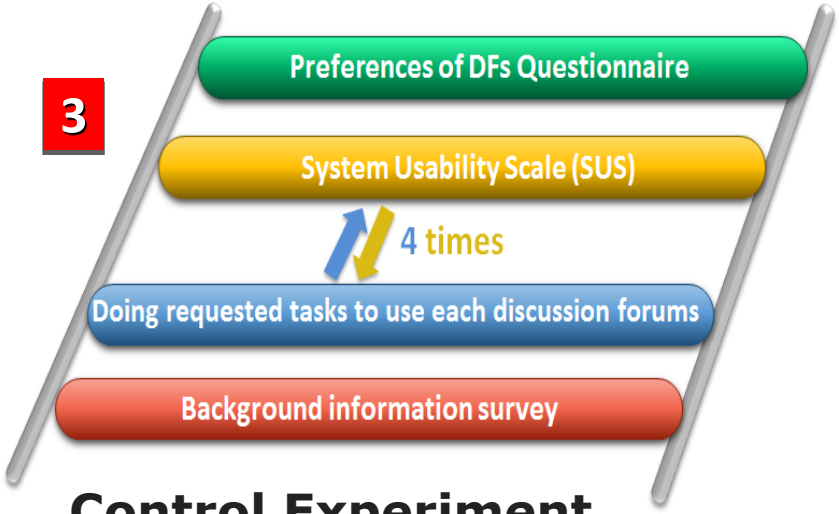


Solution

1

2

3

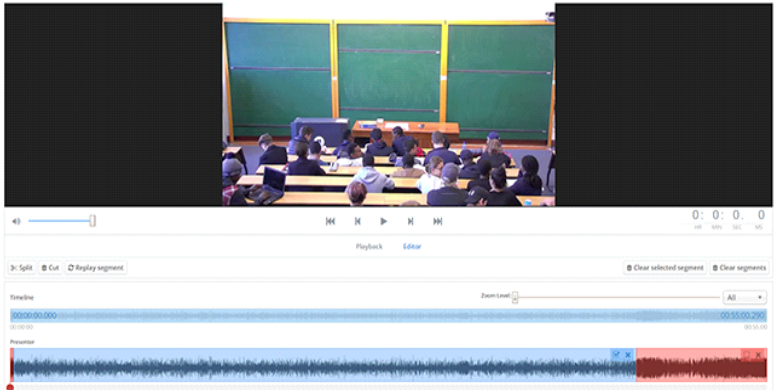


Control Experiment

MSc student: Maryam Almukhaylid

Investigating audio classification for automating the trimming of videos in Opencast

- Key Ideas
  - Classify audio into respective classes
  - Use classes to tag respective video frames for trimming
- Status
  - Training classifier
- MIT student: Devan Govender



Applications of ICT4D

- Blessing Ojeme (PhD)
- Sarah Brittan (M.CS)
- Josiah Chavula (PhD, submitted!)
- Haji Ali Haji (PhD, submitted!)
- Jessica Mantel (M.IT, submitted!)

Multi-stakeholder design of mobile micro-enterprise financial management system

- Key Ideas
  - Design mobile banking system for micro-entrepreneurs with users/NGO
  - Multi-stakeholder design (before co-design ideas)
- Status
  - Writing up thesis
- MSc student: Sarah Brittan

questions, comments, ...

http://dl.cs.uct.ac.za/

http://www.meteck.org/keen/index.html

enkosi
hamba kakuhle
thank you and go well